**SQL Analytics Project: Step-by-Step Journey in Simple Language**

**Dataset Used:** CollegeCourses

### Step 1: Getting to Know the Data

* **Procedure:** I opened the dataset and explored its structure, checking what each column represents and how the data is organized.
* **Action:** Reviewed the dataset manually and took notes of important fields.
* **Analysis:** Understood the types of insights I could get from this data.
* **Output:** A list of useful columns for analysis, like CollegeName, CourseCategory, Duration, etc.

### Step 2: How Many Unique Colleges Are There?

* **Procedure:** Used SQL’s COUNT function with DISTINCT to get the total unique college names.
* **Action:** Ran: SELECT COUNT(DISTINCT CollegeName) FROM CollegeCourses;
* **Analysis:** Checked for repeated names and ensured we’re only counting unique entries.
* **Output:** Total number of different colleges.

### Step 3: Which Districts Offer the Most Professional Courses?

* **Procedure:** Filtered rows where IsProfessional is ‘Professional Course’, grouped by district, and counted colleges.
* **Action:** Used GROUP BY, WHERE, COUNT(DISTINCT).
* **Analysis:** Found out where professional education is most concentrated.
* **Output:** Top districts offering professional courses.

### Step 4: University with the Most Colleges

* **Procedure:** Grouped data by university and counted unique colleges.
* **Action:** Used: GROUP BY University and COUNT(DISTINCT CollegeName).
* **Analysis:** Identified universities with the largest network.
* **Output:** University with the highest number of colleges.

### Step 5: Courses by Category

* **Procedure:** Grouped the records by CourseCategory and counted the rows.
* **Action:** Used: SELECT CourseCategory, COUNT(\*) FROM CollegeCourses GROUP BY CourseCategory;
* **Analysis:** Checked the balance among UG, PG, and Diploma courses.
* **Output:** Number of courses in each category.

### Step 6: Districts That Offer UG, PG, and Diploma

* **Procedure:** Filtered CourseCategory and grouped by District, checking if all three types exist.
* **Action:** Used HAVING with COUNT(DISTINCT CourseCategory).
* **Analysis:** Identified districts with the widest course range.
* **Output:** List of diverse academic districts.

### Step 7: Average Course Duration by Category

* **Procedure:** Grouped by CourseCategory and used AVG on Duration.
* **Action:** Wrote query with AVG(Duration).
* **Analysis:** Compared the academic effort required by each course type.
* **Output:** Average duration for UG, PG, and Diploma.

### Step 8: Most Common Courses

* **Procedure:** Counted how often each course name appeared and sorted them.
* **Action:** Used COUNT(\*) and ORDER BY DESC LIMIT 5.
* **Analysis:** Found which courses are most popular across institutions.
* **Output:** Top 5 most offered courses.

### Step 9: Are Courses Mostly Aided or Unaided?

* **Procedure:** Grouped by AidedStatus and counted records.
* **Action:** Query with GROUP BY AidedStatus.
* **Analysis:** Looked at funding trends in course offerings.
* **Output:** Count comparison of aided vs unaided courses.

### Step 10: Colleges Offering Both UG and PG

* **Procedure:** Filtered by CourseCategory, grouped by college, and checked if both UG and PG exist.
* **Action:** Used HAVING COUNT(DISTINCT CourseCategory) = 2.
* **Analysis:** Identified colleges offering a full academic journey.
* **Output:** List of colleges offering both UG and PG.

### Step 11: Types of Courses in Each District

* **Procedure:** Grouped by District and CourseType to count course types offered in each area.
* **Action:** Query with GROUP BY District, CourseType.
* **Analysis:** Compared learning modes in different regions.
* **Output:** Course type distribution by district.

### Step 12: Colleges with the Most Variety

* **Procedure:** Grouped by CollegeName and counted how many unique CourseNames each had.
* **Action:** Used COUNT(DISTINCT CourseName).
* **Analysis:** Checked which colleges have the broadest academic offerings.
* **Output:** Top colleges with course diversity.

### Step 13: Are Professional Courses Mostly Unaided?

* **Procedure:** Filtered data for professional courses and grouped by AidedStatus.
* **Action:** WHERE IsProfessional = ‘Professional Course’, then GROUP BY AidedStatus.
* **Analysis:** Studied funding pattern in professional education.
* **Output:** Count of aided vs unaided professional courses.

### Step 14: Average Course Duration per University

* **Procedure:** Grouped records by University and used AVG on Duration.
* **Action:** Query with GROUP BY University and AVG(Duration).
* **Analysis:** Compared universities by average academic load.
* **Output:** University-wise average course duration.

### Step 15: Course Category vs Professional Status

* **Procedure:** Grouped by both CourseCategory and IsProfessional.
* **Action:** Used COUNT(\*) with GROUP BY both columns.
* **Analysis:** Saw how professional orientation varies by category.
* **Output:** Count matrix of professional and non-professional across categories.

### Step 16: Colleges That Only Offer Diploma Courses

* **Procedure:** Isolated colleges that have only Diploma entries.
* **Action:** Used grouping and filtering to ensure no UG or PG courses.
* **Analysis:** Found diploma-exclusive colleges.
* **Output:** List of colleges focused only on diploma education.

### Step 17: Most Popular Course Type per District

* **Procedure:** Counted course types per district, and selected the one with the highest count.
* **Action:** Used subqueries or ranking logic.
* **Analysis:** Found preferred learning methods by district.
* **Output:** Most common course type per district.

### Step 18: Total Courses Offered by Each University

* **Procedure:** Grouped by University and counted unique CourseNames.
* **Action:** Used COUNT(DISTINCT CourseName).
* **Analysis:** Measured academic richness by university.
* **Output:** Total courses per university.

### Step 19: Percentage of Professional Courses Overall

* **Procedure:** Calculated total courses and then professional ones. Used basic math to get the percentage.
* **Action:** Used conditional COUNT and arithmetic.
* **Analysis:** Assessed the emphasis on professional education.
* **Output:** Overall percentage of professional courses.

### Step 20: Districts Dominated by Unaided Courses

* **Procedure:** Grouped by District and AidedStatus, compared counts to find where unaided outnumber aided.
* **Action:** Used conditional logic or HAVING clause.
* **Analysis:** Identified regions with more private education.
* **Output:** Districts with majority unaided courses.

### Final Thoughts:

This entire project helped me understand how SQL can be used to explore real-world educational data. I discovered trends in course types, duration, funding and regional focus. Writing each query step-by-step, interpreting the result and analyzing trends gave me a practical understanding of how powerful SQL is for data analysis. It was a great learning experience in turning raw data into meaningful insights.